DOCUMENT RESUME

ED 163 921 IR 006 454

TITLE	An Information Storage and Retrieval System for Biological and Geological Data. Final Report.
INSTITUTION	Smithsonian Institution, Washington, D.C.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
PUB DATE	Jun 78
GRANT	OEG-1-7-071159-4425
NOTE	54p.
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	Automation; Bibliographies; Biology; *Computer Oriented Programs; Geology; Information Processing; *Information Retrieval; *Information Systems; Interinstitutional Cooperation; Methods; *Museums; *Natural Sciences
IDENTIFIERS	*Smithsonian Information Retrieval System

ABSTRACT

This report describes the Smithsonian Institution Information Retrieval (SIIR) System, a computerized information management system for natural history museums, as well as computer developments subsequent to SIIR, and the SELGEM system which evolved from SIIR and discusses their application to data processing and collection management in museums. The methodology used in defining the scope of SIIR, the basic problems in preparing data for automation, and the definition and design of the files are covered. SELGEM, a generalized system for information storage, management, and retrieval, especially suited for collection management in museums, is an outgrowth of the SIIR system. It is used in several hundred applications by nearly all bureaus of the Smithsonian Institution. Procedures for SELGEM file updating, retrieval, sorting, editing, thesaurus processing, and indexing are discussed. A partial listing of SELGEM users at other museums and institutions is included, along with users and uses of SELGEM at the Smithsonian. Some of the more important publications relating to the system's development, application, description, and products are also listed. (Author/MBR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163921

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FINAL REPORT

PROJECT NO. 7-1159

GRANT NO. OEG-1-7-071159-4425

AN INFORMATION STORAGE AND RETRIEVAL SYSTEM FOR BIOLOGICAL AND GEOLOGICAL DATA

Smithsonian Institution
Washington, D. C.

June 1978

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

IR006454

CONTENTS

		Page
I.	Introduction	1
II.	Design of the SIIR System	2
III.	Further Development of the System . . .	14
IV.	The SELGEM System	15
V.	A Partial Index to SELGEM Utilization at Other Museums and Institutions . . .	26
VI.	Use of the SELGEM System at the Smithsonian Institution	32
VII.	Acknowledgments for Development of the SIIR/SELGEM Sytem	42
VIII.	Bibliography	44

I. INTRODUCTION

Today museum literature is replete with articles on how data may be collected, stored, processed, and retrieved more rapidly and economically than ever before because of the currently available facilities in information science. These facilities are changing the management operations of museums throughout the world. For example, over 180 university affiliated museums now have on-campus computers of sufficient capacity to support their collections management activities; and hundreds of other such institutions have access to nearby computers. Moreover, since 1970, 50 to 100 museums have installed computer-assisted systems for information management. Computer applications now number in the thousands throughout various museums, educational, and research institutions.

As one might expect, museum administrators are supplementing traditional methods of management with computerized methods and are relying more and more on the control, accuracy, and productivity obtainable from skillfully engineered computerized systems. While there is much that is automatic about automatic data processing, prior to the project discussed in this report, there was nothing automatic about adapting computer technology to the needs of museums.

In 1966, Dr. Donald Squires, Mr. Reginald Creighton, and Mr. Nickolas Suszynski designed a plan and later (1967) obtained a grant from the Office of Education, Department of Health, Education, and Welfare to develop an information processing system for museum specimens.

This report describes the results of this grant, the Smithsonian Institution Information Retrieval (SIIR) System one of the first general purpose, computerized information management systems for museums. It discusses subsequent computer developments, the SELGEM system, and their application to data processing and collection management in museums.

II. Design of the SIIR System

This section presents the basic concepts that guided the development of the Smithsonian Institution Information Retrieval (SIIR) System. It includes the methods used in defining the scope of the system, the basic problems in preparing the data for automation, and the definition and design of the files.

Procedures

The initial approach to the design of an information storage and retrieval system for museum data was to seek detailed answers to four fundamental questions:

1. What output is desired?

2. What inputs are possible?

3. What are the resources and constraints in terms of time, cost, and technology?

4. What method may be employed and what problems must be overcome so that the possible inputs may be collected and turned into the desired output in response to user's demands?

In an attempt to answer the primary question of output, scientists at the National Museum of Natural History were asked to list interrogations of the type they would like the proposed system to answer. The typical query took the following form:

> "PRESENT ALL DATA FOR SPECIMENS BELONGING TO A GIVEN TAXONOMIC GROUP AND COLLECTED AT GIVEN LOCALITY AND DURING A CERTAIN PERIOD OF TIME."

Most of the proposed queries appeared to be elaborations or variations of this general theme.

The next step in systems design was to identify the data that could be obtained and placed into the system. Data input to the system was a compromise between a vast set of possible input data and a set which was practical within the scope of the pilot system. Inputs encompass 23 to 143 discrete items of data per specimen record.

Having defined what was to be taken out of the system and what was to be placed into the system, we turned to the resources and constraints of time, cost, and technology. By

the time funding for the project was available, we had already developed a rather comprehensive systems design. Next we proceeded to identify the major problems in methodology.

Man-machine interface was of major concern. Everything possible had to be done to facilitate communication between the users, who may or may not be computer-oriented, and a rather elaborate storage and retrieval system. To achieve this goal, the following basic tenets were outlined to guide in the development of the system:

1. All formated data must be rendered searchable.

2. It should not be necessary for the system user to abbreviate his descriptors or translate them into code.

3. Query and retrieval must be simple enough to minimize user training, yet responsive enough to meet continuing time and cost-effectiveness constraints.

The Basic Problems

Three major constraints in designing the system were: the need to insure that the system be able to process a very large data base; the need to process scientific nomenclature, (the classification system used to describe the members of the animal, plant, and mineral kingdoms); and the need to process a wide variety of geographic locality terms used in the last 180 years to describe global locations.

The present holdings of the Smithsonian Institution, National Museum of Natural History are in excess of 70,000,000 specimens. Holdings in all United States natural history museums must total several hundred million specimens.

There are about 1,500,000 kinds of plants, animals, rocks, and minerals known to science. Many species may have synonyms in addition to their accepted or correct name. The number of synonyms for each species extends from none to sixty or more. In addition there are higher taxonomic groups such as class, order or family which must be considered. The total universe of names is very large and almost certainly exceeds ten million names.

In addition to problems stemming from the sheer bulk of taxonomic names, additional problem areas were defined. Two

of these were related; namely, the need to establish a structure for relating taxonomic levels, and the need to provide a method for relating synonyms so that the could respond at varying levels. These problems were resolved by establishing a hierarchical numerical code which we call numericlature.

The numericlature code provides the interface between the level of taxonomic name used in the data base and the level desired for retrieval. Specimen data records carry taxonomic names at a species level, but requests for data retrieval may cite the name of a higher taxonomic group. For example: "Retrieve all birds of family X." Therefore, taxonomic names must be structured into a hierarchy of taxonomic groups to meet anticipated query criteria. To achieve this, a system of codification based on numericlature, was adopted; this system was extended to handle synonymy. Detailed methodology was devised for computerization of numericlature as a name surrogate, synonymy rectifier, and taxonomic sequencing device. The taxonomic hierarchy used by the system is conceived of as a working tool. The structure was basically a data processing convenience and was not intended to be the definitive systematic scheme. The numericlature code portrays the following levels in this sequence: kingdom, phylum, class, order, family, genus, specific denotation, and synonymy identification.

Two nomenclature thesauri were designed to accommodate processing both by taxonomic name and by taxonomic sequence. This nomenclature-numericlature interface provided a structured hierarchy and also related synonyms. The numericlature code was automatically added to each specimen record, provided access to data through any synonym. In addition, an automatic history of synonymy for each taxonomic group and a printed listing in taxonomic sequence were easily obtained.

The problem of geographic locations was a major consideration because there were no standards for reporting the place collected and because the system needed to be able to produce all records for any arbitrarily designated area upon request. In some cases, the latitude-longitude coordinates of the collecting locality were recorded. Usually, however, a political or geographical name was the only expression of global location provided. If the system were asked to respond with all data collected in a certain global area, it needed to respond with all data for the area regardless of the manner in which the collector designated the location.

Due to the disparate forms of notation used in museum record keeping and the lack of conventions for reporting field data, it was decided that a common denominator coding system would be required to interrelate data and assure its effective recall. This code had to satisfy the following requirements:

1. It needed to be versatile. No boundaries can be imposed on retrieval, because in projecting forward, one cannot predict the global locations that will be of interest to an investigator.

2. The code had to facilitate retrieval regardless of the size or shape of an area.

3. The code had to be functionally relatable to the various reporting techniques as well as to effective data retrieval techniques.

4. It had to be compact enough to minimize the penalty created by storage of the code itself as a part of each data record.

The Global Reference Code interrelated the various reporting techniques to provide a comprehensive retrieval capability. If a verbal expression of locality was all that appeared on the specimen record, the system checked a thesaurus of previously encountered geographic expressions for a precedent to enable automatic processing of the record. Latitude-longitude coordinates from specimen records were translated directly into the appropriate global reference code via mathematical transformation functions.

Areas of virtually any size or shape may be delimited for data retrieval. These areas were defined by latitude-longitude points which, when connected by straight lines, bounded the area of interest.

In addition to the problems of taxonomy and geographic locality, other problems had to be resolved. The translation of diverse linear units into standard expressions had to be detailed for nearly all quantitative parameters to provide a basis for retrieval.

Traditional methods of manual data processing did not equip a museum with the ability to access its data according to changing criteria of selection. It was felt, therefore, that the system needed to provide storage for specimen data with complete descriptions of the physical characteristics of collected organisms and the related facts of location,

time, and event. This concept of evolutionary development and refinement of the system required that the data base be flexible, subject to content modification, and certainly content growth with minimal disruption to the system itself. Another consideration was the structural compatibility of the data base with information received by the Smithsonian from other institutions. In addition, the data base had to be structured so that the various programs utilizing and maintaining it would not be overburdened by complex procedures.

Inter-machine compatibility was an objective to facilitate potential implementation of the system at various institutions throughout the community. This would permit preservation of the system despite changes in equipment. The desire for equipment independence led to the selection of COBOL as the programming language and the use of standard software packages within the system, particularly in the area of file support.

Cost was one of the factors in the selection of data storage devices for the system. A combination of tape and disk was selected. Disk residence was used for the index files to provide ease of access and speed in query processing. Tape files were used to contain the data base because they were flexible and less expensive for volume storage.

File organizations and access methods in the disk portion of the system varied according to the need. They included the use of sequential, partioned-sequential, indexed-sequential and direct access files. Two direct access methods were used. A key transformation and randomization technique applied to alphabetic expressions was used to gain access to the Location Query Thesaurus and the Nomenclature Input Thesaurus. Actual addresses were chained to taxonomic names for access to the Abstract Search Index at varying levels of synonymy and hierarchy.

The structure and organization of the files within the system reflects the constraints of time and cost. Once the content of the major files in the system was defined and solutions found for the major problems of taxonomic structure and geographic location, details of query and retrieval were outlined. Creation and validation of the data base added another restriction. Two levels of storage were used for specimen data: a Work-In-Process level to hold intitial inputs where they were edited, verified, and cross-indexed, and a data base level where records that have passed inspection reside so as not to clutter the

Work-In-Process review file.

Early attempts to process by geographic locality revealed that information was needed in two forms. The Location Query Thesaurus was output oriented. It contained definitions of geographic names used in the query to preclude the need to define a territory already described. On the other hand, the Location Input Index Thesaurus had to function during input, by "remembering" geographic names found on input records together with their corresponding global references codes.

The development of the two nomenclature thesauri satisfied two needs within the system. The Nomenclature Output Thesaurus was used to relate nomenclatural data to specific specimen data while the Nomenclature Input Thesaurus was used for the identification of taxonomic names used on specimen records or queries so that data could be verified, stored, and retrieved by name.

Two additional indexes became vital to the system. The Redundancy Control Index contained the catalog number of each record in data base to prevent the duplicate entry of specimen records. The Abstract Search Index contained a condensation of each specimen record which in turn identified each data base record.

Specimen input data was collected on punched paper tape as a by-product of normal cataloging procedures for museum specimens. Since paper tape input comes from various manufacturers' equipment, multiple translation tables were needed to correlate the various code configurations.

As was mentioned above, the system had two levels of specimen data storage: a data base; and a temporary staging area for recent inputs - the Work-In-Process storage - which accumulated input data for review and correction. New inputs were checked by the computer and presented to the submitter for review until both machine and human audit had determined they were sufficiently accurate to be released to the data base. Validated data, released from Work-In-Process, was subjected to compression and indexing to optimize retrievability before it was added to the Specimen Data Base. Compression involved the setting of bits to record the varying conditions and characteristics of specimen data records.

For the sake of flexibility the data base was designed around a modular record concept. This offered a good compromise between fixed length records and variable length

records. The latter were compact, but complex in definition and cumbersome in processing, while the former provided efficient processing, but are space consuming. According to the modular record concept, a variable length record was carried in the system as a record comprised of a varying number of fixed length modules. Longer records had more modules while shorter records had fewer modules. Within a module, which was of a fixed standard length, data was identified by relative location. Header modules contain specimen identification information. These were followed by trailer modules containing unique detailed information about the specimen. All header modules had the same format. The number and format of the trailer modules varied depending upon the type of specimen and the amount of information available. Expansion of information could be achieved by the addition of trailer modules. The entire record was of variable length yet was some multiple of the module size.

Specimen records resided for at least one processing cycle in temporary Work-In-Process storage where they were reviewed and validated. At this stage records could be easily modified since no thesaurus posting had occurred. Work-In-Process input records, referred to as transactions, could be new entries or they could be modifications to previous accessions still contained in Work-In-Process storage. A record could be modified by changing one or more of its fields, or it could be deleted entirely.

All paper tape input records were converted into Work-In-Process transaction records, which were 3,600 characters in length with a modular format consistent with the format of data base records. There were 50 modules and all but one were identical in format and content with the data base modules. The exception was a sentinal module which carries error indicators which were set by various edits and tests while the record resided in Work-In-Process storage. It was discarded along with all empty modules when the record entered the data base.

The system indexed specimen records by geography as they were received into the system, identifying the area on the surface of the globe in which the specimen was collected and affixing to the record the appropriate global reference code. If coordinates of latitude-longitude were given on the specimen record, the transformation was automatic. If coordinates were not given, a thesaurus of verbal expressions was searched for verbiage matching that given in the specimen record. If matching verbiage was found and the thesaurus item contained a global reference code for the area, the global reference code was inserted into the

specimen record. If the thesaurus did not contain a global reference code, the specimen record was flagged and remained in Work-In-Process until either the specimen record was corrected or the thesaurus item was updated. If matching verbiage was not found, the verbiage of the specimen record was considered a candidate for the thesaurus and the specimen record was flagged pending a decision.

Initially, verbal expression of country, state, etc., were transformed to latitude-longitude coordinates. Since the verbal expression could occur in many specimen records, the entry was retained in the thesaurus for future reference. As the thesaurus builds up, fewer and fewer verbal expressions would require human intervention.

Indices were maintained to provide reference within the system components and to insure that only validated data was permitted to enter the data base. The cross-references included two nomenclature thesauri, two locality thesauri, the Redundancy Control Index, and the Abstract Search Index.

The Nomenclature Input Thesaurus permited the system to assign a numericlature code to each specimen record as it entered the system. The numericlature code superseded the taxonomic name within the system and was used to resolve synonyms and to provide hierarchical linkage and taxonomic sequence. The taxonomic name found on a specimen record had to be present in the thesaurus to be considered valid. When found, the appropriate numericlature code was assigned.

Any name in the Nomenclature Input Thesaurus could be assessed directly. The verbal expression of a taxonomic name was translated into a specific address by a process that converted letters to digits so that the number became the disk address. This method provided a random distribution of addresses and names in the thesaurus.

The Nomenclature Output Thesaurus contained a copy of all records in the Nomenclature Input Thesaurus, but the records were kept in hierarchical sequence by numericlature, and could be assessed directly if the numericlature was known. The Nomenclature Output Thesaurus served as the control of taxonomic sequencing and was the hub for adjustments to the Specimen Data Base, to the Abstract Search Index, and to the Nomenclature Input Thesaurus. The Abstract Search Index contained a coded condensation of each specimen record in the data base. Data conditions were represented by "data availability indicators." The data availability indicators consist of 240 bits (binary digits) contained in 40 alphanumeric characters. A bit set to one

(on) indicated that the data base record contained data pertinent to the condition or category represented by the bit.

The abstract record was automatically created after an incoming specimen record had been reviewed and validated. It was placed in the Abstract Search Index where it was maintained in taxonomic sequence.

The Redundancy Control Index prevented duplicate specimen records. A compound field of numeriClature-catalog number was carried in the index to identify specimen records in the system. Duplicate records were not permitted to enter the data base.

The Location Query Thesaurus provided territorial definitions of area in the form of polygons. It transformed verbal references from the query into expressions that could be tested mathematically. This thesaurus was accessed directly, once a randomization algorithm had been applied to the letters which spell territory name. Each entry on the file contained the territory or area name and the global reference code for each point on the polygon needed to describe the area. These global reference codes replaced the locality statements entered on the query for specimen record testing.

The Data Flow

Five subsystems comprised the system; each functioned on a cyclic basis.

The Nomenclature Input Subsystem facilitated the entry of taxonomic names and groups. Taxonomic names were written on data collection worksheets. Each line of the worksheet was keypunched. The keypunched data was edited and reviewed. When correct, a numeriClature code was assigned to each taxonomic name. Each record was then added to the Nomenclature Output Thesaurus and the Nomenclature Input Thesaurus.

The Specimen Input Subsystem facilitated the entry of specimen records into the Work-In-Process file. Once initial checks and corrections had been made, each specimen record received into the system had its taxonomic name verified by the Nomenclature Input Thesaurus. The corresponding numeriClature code was inserted into the specimen record to serve both as an identifier and as a sequence key.

In the Specimen Release Subsystem the specimen records

were passed from the Work-In-Process file to the Specimen Data Base file. The same data was also compressed into abstract records which were added to the Abstract Search Index File. When new specimen records had been placed into the data base and the Abstract Search Index, the Nomenclature Input Thesaurus was posted with the address of the data so that it could be recalled by taxonomic group name. The Nomenclature Input Thesaurus was also posted with new names and groups that had entered the system. The address of the first specimen abstract record for each taxonomic group on the Abstract Search Index was also placed in the records on the Nomenclature Output Thesaurus that pertained to that group.

The Location Query Thesaurus Subsystem was used in the query process to provide territorial definitions in the form of polygons.

The Query and Retrieval Subsystem validated and edited the query statements, located the abstracted data, retrieved the selected specimen data and presented the output in the requested sequence.

Five basic parameters constituted a query: Taxonomic name, geographic locality, time-span, general conditions of retrieval characterized by boolean logic and data availability indicators, and specific arguments for text search.

The key for locating specimen abstracts was obtained through the name of the taxonomic group. When a taxonomic name was located on the Nomenclature Input Thesaurus, two significant pieces of information were retrieved: the numeriClature code for the taxonomic group and the address that located the first specimen abstract record for the group in the Abstract Search Index. Search was, therefore, precluded if the taxonomic name could not be located on the thesaurus.

The abstract records were then examined starting with the address obtained from the Nomenclature Input Thesaurus and continuing until a comparison of numeriClature codes indicated that the search was completed. Abstract records contained the data necessary for testing time-span, locality, the miscellaneous general conditions of data, and the data availability indicators. The numeriClature and the catalog number from the selected abstract records provided quick identification of the data base records that qualified for detailed inspection in compliance with the final query criteria for text search. During the final phase of retrieval the selected specimen records were written onto a

magnetic tape in preparation for printing or for executing other output options.

Results and Conclusions of the SIIR System

The System provided the first interdisciplinary data base for specimen related data in the field of natural history. The principal objective of the system was to utilize museum collections more efficiently in education and research by furnishing rapid, in-depth access to the information content of specimen records and related bibliography. This objective has been met.

The SIIR system, with its global reference and nomenclature indices combined with a binary representation of abstracted data conditions, offered a direct solution to many of the problems stemming from the ever-increasing collections and specimen data. The benefits of automated cataloging and review procedures also served to enhance the information content of museum records.

The system provided the scientific investigator with an ability to direct a broad spectrum of questions to an interdisciplinary data base. It remained only for the museum to maintain and improve the system and enter data so that the scientists could ask their questions.

Data, once entered into the system, was never lost. Nevertheless, some deterioration was possible due to human factors. A human review of the Work-In-Process storage and of the location input index Thesaurus was mandatory each time new inputs were transferred from Work-In-Process to the Specimen Data Base. Many automatic safeguards were provided, but no automated system, regardless of how well-conceived, could replace human judgment. At best, computerized edit and audit programs could check data for validity only within broad parameters. An investigator who adds to the data base had to be responsible for the validity of all data under his control.

One of the most significant points learned in the process of developing this system was that a viable information system is an integration of many elements, including computer hardware and software, data, and people. If any one of these is lacking for any extended period of time, the system will be severely damaged. Data must continue to be entered. People must continue to interact. Continuous programming is required to keep pace with software revisions and to keep documentation up-to-date. User criticism should feed improvements back into the system. The museum must

continually investigate and incorporate new techniques to keep the system on a par with current standards of the art both in subject matter and in processing standards.

There were few, if any museums or data centers that had a museum-appropriate information storage and retrieval technology prior to this project, and there was a growing gap between the few with the capability and many with only the need. The SIIR system, at project completion did not represent the ultimate in museum systems, but it was a good beginning. We had gone beyond feasibility studies and surveys into the design, implementation, and evaluation of an automated system. The development of this system may or may not represent a breakthrough in the retrieval of specimen data, but it was a pioneering effort without precedent which identified many problems and solved some of them. Those who witnessed demonstrations of the system, regarded it as having the potential to greatly increase the value of collections for research and education.

The construction of this information system was a unique venture. It was the only system that satisfied the critical mandates for possessing specimen based biological data-nomenclature, hierarchy, synonymy, and global reference. SIIR and similar systems have continued to be developed and used to support museums in the fulfillment of their information responsibilities.

III. Further Development of the System

Systems continue to grow and evolve or they soon die out. Fortunately, the SIIR system continued to grow and change.

The system was further generalized to facilitate user communication. Due to time and cost constraints, some portions of the system had been tailored to particular expectations. These portions were reprogrammed and generalized to meet broader needs. Systems analysts continued to evaluate the system's performance with respect to expanding user requirements, cost, storage allocation, timing optimization, etc.

During the last year of the SIIR project, Mr. Creighton, together with Mr. James Crockett and others at the Smithsonian's Office of Computer Services (OCS), began redevelopment of the SIIR system. This evolutionary son of SIIR is called SELGEM (SELf GEnerating Master). The purpose of the redevelopment was to rectify shortcomings of the SIIR system and to take advantage of what was learned during the develop ment and use of SIIR.

This improved system, SELGEM, has proven highly successful and of interest to many. Over 700 museum professionals have requested information concerning the system. The computer program of SIIR-become-SELGEM and its documentation have been made available to over 80 cultural institutions throughout the world. These institutions communicate with each other through an organization called Museum Exchange for Systems Help (MESH).

SIIR/SELGEM, then, is a product of ten consecutive years of research and experience with museum data processing. Throughout the years various museum and computer professionals made diverse and invaluable contributions to SELGEM's design and function. Reginald Creighton designed the central structure of the system. The further design and development of SELGEM has been a team effort involving many participants, notably Mr. James Crockett and diverse programmers and system users (See the Acknowledgments Section).

IV. The SELGEM System

The SELGEM system is a generalized system for information storage, management, and retrieval especially suited for collection management in museums and many other applications. SELGEM is used in several hundred applications by nearly all bureaus of the Smithsonian Institution.

To use the SELGEM system the user selects the computer programs and options best suited to his current needs and combines them together by submitting a sequence of control cards to the computer operation facility. By combining the options of several general purpose programs, the user circumvents the costly and time consuming requirement to have numerous, temporary, special-purpose programs written. The system offers the ability to adjust a record's length, content, and format on a spontaneous basis and accommodates unanticipated data without any reprogramming. Modularized components provide a choice, according to budget and need among various levels of sophistication. Because the SELGEM system is generalized and modularized no capital investment is necessary for a new user.

SELGEM provides a generalized method to search, and to print selected outputs, and it also provides generalized routines for such functions as updating, sorting, report writing, editing, thesaurus processing, as well as for bibliographic and collection indexing. The system is operated in batch-processing mode with the retrieval program accessible by remote terminal.

Another feature of interest is SELGEM's mobility. It is written in COBOL language, making it highly equipment independent. It can be installed by journeyman programmers on any computer with four tape units, a COBOL language compiler and 64,000 characters of main storage. No larger capital investment is required.

How SELGEM Works

1. The File Update Program

The notion that only pure and complete collection records should be placed into a computer system defeats one of the major benefits of computerization - the use of the computer as an updating device.

SELGEM facilitates the computerization of in complete records by affording the user the ability to input, modify, and delete data, one field at a

time. Beyond this, it provides various options for selective modification and text editing and because of its unique data-base conventions, SELGEM also permits a distribution of labor with respect to input; that is, a person at one level of skill can enter data as it appears on existing records, while a person at another skill level can do review and clean-up work. On a bibliography project, for example, one person can record the author, title, date, publisher, and collation while another, perhaps at a different time, can supply subject terms or key-words for indexing.

SELGEM's update procedure is simple to comprehend and to execute. Before getting into procedure, a review of the data-base record format will be helpful. The fixed-length units, or modules, of information that make up data-base records are comprised of 64-character positions for data and 13-character positions for identification and sequence control. The latter is composed of an 8-digit catalog or serial number to identify and sequence records; a 3-digit category number identify fields and to sequence fields within a record; and a 2-digit line number to sequence the lines that comprise fields exceeding 64 characters of data. Thus, an item of data (field) containing an expression, say, 128 characters in length, would be carried by the system as two lines of 64 characters each (labeled 01 and 02).

The category code, left to the discretion of the user, identifies the data (for example, 070 might identify an author, 090 might identify a geographic locality). The specimen record contrived below for catalog number 166886 illustrate the technique. Note that category 130 required two lines.

```
00166886  030  01  ECTOMARIA

00166886  050  01  MARYLANDICA

00166886  070  01  CLARKE AND SWARTZ, 1933

00166886  090  01  UNITED STATES

00166886  100  01  MARYLAND

00166886  110  01  SIDELINE HILL CREEK
```

```
00166886  120  01  2.5 MI. ABOVE MOUTH

00166686  130  01  NOW IS THE HOUR..AID OF

00166886  130  02  THEIR COUNTRY
```

The master-file is maintained in an ordered sequence by serial number, category number, and line number. Note that one module fits onto one punched card with three columns remaining for special purposes. On magnetic tape trailing blank-characters are deleted to conserve space and processing time.

Transactions to update a master file may be prepared using several data entry techniques, including punch-paper-tape, optical-character-recognition (OCR), punched cards (in several formats), key-to-disk systems, and magnetic tape from special programs or other systems. Several of these techniques require preprocessing programs to create standard transaction records, which are based on 80-character punch card format. If the transactions are prepared on punch cards, preprocessing is not required; this format illustrates the standard transaction format. The card format divides into two sections: the identification and control portion in columns 1 through 14, and the data portion in columns 16 through 79. The identification and control section is divided into the serial number, columns 1 through 8; the transaction code, column 9, the category number, columns 10 through 12; and the line number, columns 13 and 14. The use of columns 15 and 80 are reserved for a special function of a change transaction. The transaction code identifies the type of transaction such as add, delete or change. Preprocessing programs for other data entry methods output transaction records in the above format.

The efficient production of machine-readable museum records has long been recognized as a problem of major importance. Squires (1966) and Manning (1969) comment on the rapid growth rate of collections and describe the use of punched-paper-tape equipment and techniques to remedy the problem. The use of punched-paper-tape as an input media is especially desirable when the tape

can be obtained as a nearly gratuitous by-product of normal typing, or where equipment and personnel are already available, otherwise a strong case can be made to use other input devices and media.

The paper tape input procedure is simple, with few restrictions placed upon the data preparator. It requires three unique symbols to represent: 1) beginning-of-record, 2) beginning-of-category, and 3) end-of-record. A parameter card identifies those symbols used for each file. A category of data is recorded as follows: the beginning-of-category symbol is typed (or emitted by a control program loop), followed by a 3-digit category number to identify the data which follows. (Line numbers are assigned automatically when the punched-paper-tape is processed by the input program.) The data entered may be of variable length extending from line to line on the typewriter. It is terminated by the next beginning-of-category indicator.

A simple error-recovery procedure is provided. If an error has been made, the record may be started over by omitting the end-of-record character. During processing, any record lacking an end-of-record character is dropped.

In addition to identifying characters employed to separate categories and records, the data preparator must also provide the category number of the data field that contains the "serial number." A specimen's catalog number is frequently used as "serial number," but any system that uniquely identifies each record, is acceptable. Paper tape in any code configuration may be processed and may have originated on various devices made by Friden, Dura, SCM, Teletype, etc.

2. Retrieval

Briefly, the procedure for retrieval using SELGEM requires the following steps. A record is tested for the presence (or absence) of a given data word among the words in a particular category. A designated indicator is turned on if the word is found. (Up to 99 indicators may be set for a record.) Then the pattern of indicators set for a record is compared with the pattern specified by the user. If the patterns are the same, then one

of four actions will be preformed. The four types of action are: output, print, tally or replace data.

Data Evaluation

Any word in any category may be specified for search. A word is defined as any string of characters bounded by spaces. Thus, XYZ is one word, but X Y Z are considered three words. The word (or value) that is entered may contain letters, numbers, or both. Punctuation marks (comma, slash, colon, semi-colon) terminates a word, but the period does not. The period is used as a decimal point or for retrieving abbreviations.

Two types of numbers are recognized by the retrieval programs: real numbers, which have a decimal point, and integers, which have no decimal point. If the search is made for integer data, a number of magnitude 9999999999999 can be specified. But, if the search is concerned with real data, decimal alignment takes place, and the largest possible number is 9999999.999999. Comparisons between real and integer data produce unpredictable results. Hence, if the data item in the file contains a decimal point, then the value entered on the worksheet must contain a decimal point.

Category numbers, which identify the various classes of data, are entered as 001 and 010. Word number refers to the relative position of the word in the category. To search all words of a category, enter 00 as the word number.

EXAMPLE A:

If the user wants to know every time the word Juniper appears as the first word in an entry in the category vegetation types, 031. In the DATA EVALUATION section he writes:

TURN ON INDICATOR 01, IF CATEGORY 031, WORD 01, IS EQUAL TO JUNIPER.

EXAMPLE B:

If the user wants to know, in the date collected category (014), all items collected through 1957. Where date is recorded as (day, month, year - 03 DEC 1967), the year is the third word. He writes: TURN ON INDICATOR 02, IF CATEGORY 014, WORD 03, IS NOT GREATER THAN 1957.

Indicator Evaluation

The indicator evaluation statements test the relationship of the data element in a record to the argument specified by the user in the data evaluation statement and specify the action to be taken when the desired relationship is found. The indicator is "ON," if the data evaluation criterion is met; "OFF," if it is not. The action number is arbitrarily assigned and provides the link to the definition of the action.

One hundred and fifty indicator evaluation tests can be entered in one query and 30 actions. Different indicator tests can address the same action number and an indicator can be tested more than once.

EXAMPLE:

To find the number of records belonging to a certain genus the indicator test would be a simple:

IF INDICATOR 01 IS ON THEN PERFORM ACTION 01.

If the total of other genera on the file is also wanted, add:

IF INDICATOR 01 IS OFF THEN PERFORM ACTION 02.

Since all data evaluation is completed before the indicators are tested, relationships among data conditions can be developed using the "AND" and "OR." This allows us to further limit the previous example to the species level.

EXAMPLE:

In this case the generic name is associated with indicator 01 and the species name with indicator 02. The indicator test becomes compound:

IF INDICATOR 01 IS ON AND INDICATOR 02 IS ON THEN PERFORM ACTION 01.

To find the presence of other species in that genus, add:

IF INDICATOR 01 IS ON AND INDICATOR 02 IS OFF THEN PERFORM ACTION 02.

Very complex relationships can be tested with the SELGEM query program by selectively combining the Boolean logic of the indicator tests with the expression tests of the data evaluations.

The Use of AND

When more than one data condition must be evaluated to determine retrieval, an indicator is specified in the DATA EVALUATION section for each data condition. The indicators are linked in the INDICATOR EVALUATION section and tested to see if the resultant "ON" or "OFF" status of the indicators reflects the desired combination of conditions.

EXAMPLE:

To illustrate the testing of indicator relationships, suppose that an airline wishes to review the records of employees to select candidates for flight service opportunities. They want persons between the ages of 40 and 65, that are less than 6 feet tall, and weigh less than 180 pounds. In terms of the computer, we consider the problem as follows:

If age is less than 40, turn on indicator 01.

If age is greater than 65, turn on indicator 02.

If height is less than 6 feet, turn on indicator 03.

If weight is less than 180 pounds, turn on indicator 04.

We test the status of the indicators in terms of:

IF INDICATOR 01 IS OFF, AND INDICATOR 02 IS OFF, AND INDICATOR 03 IS ON, AND INDICATOR 04 IS ON, THEN PERFORM ACTION NUMBER 01.

The Use of OR

The use of OR indicates the start of a new set of condition tests. The statement A and B or C is evaluated as meaning either A and B, OR, if not, then C alone. If A and B and also A and C is intended, it must be so expressed.

EXAMPLE:

Some specimens in a file have been identified under different generic names and the user wants to retrieve all records in which the genus X-us was used as a senior name for the genus Y-us. The category code for the senior name data is 050. Previous scientific names are carried in categories 400 and 410. DATA EVALUATION statements have been written to test for the particular names and to turn on indicators 01, 02, and 03. The following is the INDICATOR EVALUATION statement:

IF INDICATOR 01 IS ON AND INDICATOR 02 IS ON OR INDICATOR 01 IS ON AND INDICATOR 03 IS ON THEN PERFORM ACTION NUMBER 01.

Actions

The ACTION section defines the type of actions to be performed, if the proper indicators have been turned on, and if they have proper relationship to each other. These actions include: printing (onto paper), or copying records or selected categories from records onto tape; tallying the number of times the data meet the criteria; and replacing sequences of characters (modifying data content). Each action statement must have a unique action number corresponding to the action number in the indicator tests. Thirty actions may be specified.

3. SELGEM Report Writer

Through the report writer, the user prints reports in any format desired. A short deck of punched parameter cards specifies the content and format of the report. Partial lines, complete lines, categories, or entire records may be printed in any sequence.

To the contents of each field selected for printing, the user may attach a constant prefix or suffix. For example, the express "Depth" followed by a blank character may be prefixed to a depth measurement with the expression "Meters" added as a suffix.

The parameter card that identifies the data to be printed also specifies the location on the printed page where the data must appear. A field to determine carriage return and line spacing is included.

Data may be carried in the SELGEM master-file in a codified representation and decoded by the report writer into plain text for printout. To decode, the coded data items in a category are looked up in a list (supplied on punched cards) and the list entry is substituted for the data item. Data items not found in the list are written out untranslated. The width and depth (number of lines) of a printed page may also be specified. Data may be printed in a tabular format or in a free flowing paragraph format, or a mixture of both. Provision is made for multiple line report headings and page numbering. Data items may be printed with leading zeros removed.

4. Sorting

The order in which records are stored in a system is frequently not the same order in which a user wishes to see them on output. SELGEM has a sort package (two programs) which can sort any data base master file into the desired sequence. The user need only specify the category numbers, line numbers, starting character position and length of the data to be sorted. The output from the retrieval program and the indexing program is frequently sorted before being printed by the report writer.

5. Editing

The ability to test data for validity by comparing it to various norms, ranges, formats, and conditions is often critical to reaching the goals for which the data was collected. For example, if retrieval or plotting, based on a variable such as month of the year, is anticipated, then a convention for recording months must be employed and adherence to the convention must be verified. Suppose each month were represented by three letters, e.g., JAN for January, FEB for February, etc. A computer program could compare each month submitted as data with a list of standard month abbreviations (authority list) to insure that each expression of month is correct. This type of verification is one of a series referred to as "editing" and is performed by an "edit program."

Because consistency of data format is so often a prerequisite to proper data utilization, the

effectiveness of the edit program can have a significant impact on the viability of the information system. Unfortunately, edit programs are frequently highly specialized and inflexible. Typically the location of the data to be edited is defined and incorporated into the program in advance. In addition, the norms for verification of the data are also cast as a part of the program itself. When both data format and norms for verification are built into an edit program it can become so highly structured that it has utility for only one file. Proliferation of costly edit programs only channels resources away from more significant application.

To overcome the usual tendency toward a multitude of rigid edit programs, the SELGEM edit program is generalized. It employs parameter cards, introduced just prior to program execution, to specify norms and data locations. Thus, the same program can be applied to any other file of data by supplying parameter cards appropriate to that file.

6. Thesaurus Processing

Frequently, a machine readable thesaurus plays a vital role in processing. The automated thesaurus functions as follows: relates synonyms to each other, establishes equality or seniority, ascribes class membership and hierarchy to discrete items of data, validates spelling, provides surrogation, performs a host of other chores, such as posting and tally-keeping.

Data in SELGEM format can be linked to various kinds of thesaurus processing. A nomenclature thesaurus can be used to relate synonyms, to maintain taxonomic sequence, and to validate the nomenclature contained on incoming specimen records. Geography is a major problem to a natural history museum, due to the lack of a standard notation for global areas and also to the constraint and expectation that a data retrieval system be capable of extracting all pertinent information in response to a query regardless of the manner in which location is reported into the system. A global-reference code based on latitude-longitude coordinates working in conjunction with an automated thesaurus of the most

frequently occurring expressions of locality, offers a solution to this problem.

Without any requirement for a prior definition of territorial boundries, computer records containing the global-reference code can be promptly retrieved according to locality.

7. Indexing

Computer-generated subject and author indices to bibliographies, specimen records or other collections are easily obtained. A SELGEM subsystem for indexing identifies index terms and creates an index record for each term. These indices can then be sorted and printed with the SELGEM report writer.

A special print program for bibliographic indices displays data in alphabetical sequence by author. Co-authored entries are sequenced according to the name of the first author. Title and journal information is brought together to get the most on each printed line.

There are two parts to an entry in the subject index: 1) the first line which gives key-word, author, and date; and 2) subsequent lines which contain associated key-words to help convey the context of the indexed term. The index is ordered alphabetically by key-word, author, and date. Cross-reference is thereby provided to the citation-list, which is ordered by author and date.

The listings are printed in conventional publication format. On odd numbered pages, the page number prints in the right margin and on even numbered pages, the page number prints in the left margin. A one-line heading, supplied by the user, prints on odd numbered pages with a different user-supplied heading printed on even pages. "See" and "see also" entries may be used.

V. A Partial Index to SELGEM Utilization at Other Museums and Institutions

SELGEM - Users by Institutions

ACADEMY OF NATURAL SCIENCES, IBM 370-168
MOSS, DR. W. WAYNE
AMERICAN UNIVERSITY, IBM 370-145
HUME, MS. VALERIE
MCNETT, DR. CHARLES
ARIZONA STATE MUSEUM, CDC 6400
CHAFFEE, MS. HOLLY
CUSHING, MR. JAMES
MANIRE, MR. LARRY
ARIZONA STATE UNIVERSITY, UNIVAC 1110
GAINES, DR. SILVIA
SMITH, DR. L. D.
AUBURN UNIVERSITY, IBM 370
HILL, MR. EDWARD P.
BOTANICAL RESEARCH INSTITUTE, B 6700
CODD, DR. L. E.
BRIGHAM YOUNG UNIVERSITY, IBM 360-65, DEC 10
BRIMHALL, PROF. WILLIS H.
HERRIN, DR. SELBY
HODSON, MR. RICHARD B.
JOHNSON, MR. CRAIG
MATHINI, DR. RAY
PRITCHETT, DR. CLYDE
RIGBY, DR. KEITH
SAUNDERS, MR. ROBERT C.
TIPTON, DR. VERNON
BUREAU OF LAND MANAGEMENT, CDC.3100,IBM 360-65
GREEN, MR. ROBERT
CIUDAD UNIVERSITARIA, IBM
KING, DR. JAMIE LITVAK
EASTERN NEW MEXICO UNIVERSITY, IBM 360-40
IRWIN-WILLIAMS, DR. CYNTHIA
JOHNSON, PADDY C.
PIOTTER, MS. SUSAN
FIELD MUSEUM OF NATURAL HISTORY, H 2015, SI
SOLEM, DR. ALAN
TURNBULL, DR. WILLIAM D.
FLORIDA DEPT OF STATE
FRYMAN, JR., MR. FRANK B.
FLORIDA STATE MUSEUM
HUMPHREY, MR. STEPHEN R.
WEBB, DR. S. DAVID

WING, ELIZABETH S.
FOREST SERVICE, CDC 3100, UNIVAC 1108
DE BLOOIS, DR. EVAN I.
NEVES, MR. RICHARD
ONDLER, MISS LUCILLE
SAFRAN, MR. R. L.
ILLINOIS STATE UNIVERSITY
CHRISTENSEN, MR. JAY
HANSEN, MR. ARNE
IOWA STATE UNIV. OF SCIENCE & TECH.
BOWER, MR. JOHN R. F.
JOHNS HOPKINS UNIVERSITY, H 2015, SI
LOUISIANA STATE UNIVERSITY, IBM 360-65
ERICKSON, MR. RICHARD
MAINE STATE MUSEUM, IBM 370-135
DAMM, DR. ROBERT L.
KLEY, MR. RONALD J.
MUSEUM OF ALBUQUERQUE
BORHEGHYI, SUZANNE DE
MUSEUM OF NEW MEXICO, IBM 360-65 UNM
COX, MIKE
EWING, GEORGE
LANGE, MS. YVONNE
PECKMAN, STEW
REILEY, ANITA
SVERO, DEE DEE
WEBER, MR. MIKE
MUZEJSKI DOKUMENTACTIONI CENTAR
BAUER, DR. ANTUN
NATIONAL CENTER FOR HEALTH STATISTICS
GRIGSTON, MS. JEAN
HAYSLETT, MR.
ROTHWELL, MS. SANDY
SPILLANE, MR. WILLIAM M.
NATIONAL INSTITUTE OF HEALTH, H 2015, SI
CLIFFORD, MR. CARLETON M.
NATIONAL PARK SERVICE
FULGHAM, MR. TOMMY R.
LYONS, TOM
NATURE CONSERVANCY, H 2015, SI
JENKINS, DR. ROBERT E.
NEW MEXICO REGIONAL MEDICAL PROGRAM, PDP1140
PATTERSON, MR. RICHARD
OBERLIN COLLEGE, IBM 360
WANGER, PETER K.
OHIO HISTORICAL SOCIETY
ALBRECHT, MR. CARL W.
PORTSMOUTH POLYTECHNIC

NEWTON, MR. PHILIP J.
PUBLIC HEALTH SERVICE, H 2015, SI
CHAMBERLAYNE, DR. EARL C.
SMITHSONIAN ASTROPHYSICAL OBSERVATORY
WHITMAN, MR. JOHN
SMITHSONIAN INSTITUTION, H 2015, SI
ADAMS, MS. ANNE R.
AYENSU, DR. EDWARD S.
BEDINI, DR. SILVIO A.
BETTS, VIRGINIA
BLOCK, MS. JUDY
BOOTH, MS. ABIGAIL
BRIDGE, MR. DAVID
CHILD, MR. C. ALLEN
COLLIER, MR. FREDERICK J.
COWAN, DR. RICHARD S.
CREIGHTON, R. A.
CROCKETT, MR. JAMES J.
DAVIS, DR. AUDREY
DESAUTELS, DR. PAUL
DOWNEY, MS. MAUREEN E.
DURANT III, FREDERICK C.
EILERS, WILLIAM
ERWIN, LA VERNE
EVANS, DR. CLIFFORD
FITZHUGH, DR. WILLIAM W.
GAUSH, ARTHUR L.
GAUTIER, DR. GARY
GIBBS, DR. ROBERT H.
GRAMBLIN, MR. EDGAR N.
GRANT, DR. RICHARD E.
HANDLEY, DR. CHARLES
HARSHBARGER, DR. JOHN C.
HIGGINS, DR. ROBERT
HOPE, DR. DUANE W.
HORTON, MS. NIKKI
HURD, JR., DR. PAUL D.
JACKSON, DR. MELVIN H.
JOHNSTON, MR. WILLIAM R.
KERR, MR. GEORGE
KIER, DR. PORTER M.
KOVY, MR. STANLEY A.
KROMBEIN, DR. KARL V.
LANDRUM, MISS BETTY
LAWSON, MR. WILLIAM L.
LECKIE, MS. DORIS J.
LERNER, ABRAM
LYTLE, RICHARD H.

MANNING, DR. RAYMOND B.
MASON, DR. BRIAN H.
MELLO, DR. JAMES F.
MELSON, DR. WILLIAM
MERZBACH, DR. UTA C.
MORRIS, MS. MARTHA
ORGAN, DR. ROBERT M.
PACKARD, MS. PENELOPE A.
PAWSON, DR. DAVID L.
PETERSON, MS. KATHY
PIERCE, DR. JACK
PONTE, MR. VICTOR M.
PURDY, MR. ROBERT W.
QUADE, MS. ELIZABETH R.
ROBINSON, DOUGLAS J.
ROBINSON, EDWIN
ROPER, DR. CLYDE F. E.
ROSEWATER, DR. JOSEPH
SEFEKAR, MR. JOSEPH
SERIO, MS. ANNE
SETZER, DR. HENRY
SHETLER, MR. STANWYN
SMALL, MR. GREGG J.
STEPURA, MR. AUGUST
THORINGTON, DR. RICHARD
UYAMA, MR. HIROSHI
WALLER, DR. THOMAS R.
WATSON, DR. GEORGE
WOLFE, MR. ROBERT
ZUG, DR. GEORGE

SOUTH DAKOTA STATE UNIVERSITY
BALSBAUGH, JR., DR. EDWARD U.

STANFORD UNIVERSITY, IBM 360-67
GEROW, DR. BERT A.
NALL, MS. M. LEXIE

STATE OF NEW MEXICO
ELLIS, MR. FRED

STATE UNIV. OF NEW YORK, BUFFALO, CDC 6400
HURD, MR. ROY
TRUBAWITZ, MR. NEIL
WELTE, MR. JOHN

TEXAS TECH UNIVERSITY, IBM 370-145
BUTLER, III, MR. PATRICK H.
BRAY, ROBERT
CRITTENDEN, MARY
ESTES, MS. PAM
KING, DR. MARY ELIZABETH
PURKAPLE, MS. GAIL

U. S. AIR FORCE, CDC 3600 WALTER REED
JACKSON, JR., COL T. F.
U. S. DEPT OF AGRICULTURE, H 2015, SI
BURKS, DR. BARNARD D.
U. S. DEPT OF THE INTERIOR U S GEOLOGICAL SURVEY, H 2015, SI
OLIVER, JR., DR. WILLIAM A.
UNIVERSITY OF ALASKA
ALLISON, DR. CAROL
DIXON, JR., MR. E. JAMES
MUELLER, MR. GEORGE J.
UNIVERSITY OF ARIZONA, CDC 6400
BRUSCA, MR. RICHARD C.
HUMPHREY, MR. ALLEN
UNIVERSITY OF ARKANSAS, IBM
ALLEN, DR. ROBERT T.
UNIVERSITY OF CALIFORNIA, BERKELEY, CDC 6400, IBM 360
BROWNE, MR. ANTHONY L.
CHENOWETH, MR. TED
DURHAM, DR. J. W.
MANCEBO, MS. LINDA
NORICK, MR. FRANK
PECK, JR., DR. JOSEPH H.
WAHORTEN, MR. ROBERT
WALKER, MR. WAYMAN
UNIVERSITY OF CALIFORNIA, DAVIS, B 6700
COLE, MR. RONALD E.
SOMERBY, MR. RONALD E.
UNIVERSITY OF CALIFORNIA, LOS ANGELES, IBM 360-91
ERICSON, MR. JONATHON E.
UNIVERSITY OF CALIFORNIA, SANTA BARBARA, IBM 360-91/75
GLASSOW, DR. MICHAEL
GREGG, MR. GEORGE
JEFFRIES, MR. RONALD
WATTS, MS. BONNIE
UNIVERSITY OF CAPE TOWN, B 6700
GURNEY, DR. JOHN
HALL, DR. A. V.
UNIVERSITY OF COLORADO, CDC 6400
DOHRMANN, RUSSELL
RODECK, MR. HUGO
UNIVERSITY OF FLORIDA
WEBBER, E. LELAND
UNIVERSITY OF KANSAS, H 635
HOFFMANN, DR. ROBERT S.

HUMPHREY, DR. PHILIP S.
MICHENER, DR. CHARLES D.
NEUNER, DR. A. M.
ROWELL, DR. A. J.
WOLFE, DR. PAUL J.

UNIVERSITY OF MISSOURI, COLUMBIA
MARQUARDT, MR. WILLIAM H.

UNIVERSITY OF NEW MEXICO, IBM 360-65
CORDELL, DR. LINDA
CRAWFORD, MR. CLIFFORD
STRUEVER, MS. MOLLIE

UNIVERSITY OF RHODE ISLAND
PETERSON, DR. HAROLD

UNIVERSITY OF TEXAS, ARLINGTON
DARLING, MR. JOHN L.

UNIVERSITY OF TEXAS, AUSTIN, CDC 6600
WEDEL, MR. WALDO M.

UNIVERSITY OF TEXAS, EL PASO, IBM 360,
TEXAS TECH
HARRIS, DR. ARTHUR
O'LAUGHLIN, DR. THOMAS C.

UNIVERSITY OF TRONDHEIM
KLAUSEN, MR. IVAR

WALTER REED INSTITUTE OF RESEARCH, CDC 3500
CROSS, MISS
HANNEGAN, M.D., LTC MC M. W.
LEE, MR. ROBERT
MILO, MR.
ROBINSON, MS.
WENNER, 1LT PATRICIA
WILBURN, LTC.

WESTERN MICHIGAN UNIVERSITY
ANEMA, MR. SAM
JACOBS, DR. ELLEN
NEUSIUS, MR. PHILIP

VI. Use of the SELGEM System at the Smithsonian Institution

At the outset, the SELGEM System was focused on the data processing applications and problems of the natural history museum. The SELGEM System has proved so versatile, flexible, and comprehensive in the area of data management (particularly collection management), that it is now used by nearly all bureaus of the Smithsonian Institution. The System is serving the Institution well; the Institution is supporting continued development of the SELGEM System to meet the increasing and changing requirements of the world's largest museum organization.

1. National Museum of Natural History (NMNH)

The NMNH is currently the biggest user of the SELGEM System. Over 1.5 million records represented by more than 15 million data lines, are stored under the SELGEM System. These data are contained in about 250 distinct computer files, each a mini-collection for processing purposes. Some of the more important applications are:

Anthropology Department - Ethnology. An index file is nearly compiled for the entire collection, using the existing catalog-card file as the source of information. The prime objective is to use the file to help conduct a physical inventory of the collection in order to review the status of the objects and to verify the data. A current catalog file may soon be started.

Anthropology Department - Archeology. A retrospective index file has almost been completed for a portion of the collection, and procedures are being set up to build and maintain a current catalog file. Department staff are taking a leadership role in defining national objectives and data standards for computer processing of archeological collections.

Anthropology Department - Physical Anthropology. An index file has been built for the entire collection of human skeletons. The file is part of the registry of paleopathological human specimens, which provides a source of information on the micro-evolution of diseases in human populations. A bibliographic reference file is also part of the registry.

Botany Department - Type Collection. Information on over half of the museum's collection of 60,000+ type specimens of plants has been put in a retrospective/current catalog file. Upgrading and verification of data are key goals of the project. The file has been used several times to provide information to researchers.

Botany Department - Index Nominum Genericorum. Publication information for all valid genera of plants is being computerized for publication in 1979 and for information retrieval purposes. Though the file does not contain information on specific specimens, it does provide information on genera and type species which may be represented in the collection.

Botany Department - Endangered Flora. A computer file has been built of information on all species of plants which are either recommended, proposed, or officially listed as endangered and threatened in the United States. An early version of the file was used to publish the list. Specimens of some of the species on the list do occur in the herbarium.

Entomology Department - Carabid Beetle Collection. A research/current cataloging file is being built for the museum's holdings of ground beetles from areas of Central America as well as those of a local island in the Potomac River. Approximately 54,000 specimens have been entered so far. The project has served as a model for several other scientists who are now building compatible files.

Entomology Department - Catalog of Hymenoptera in America North of Mexico (reference file). Extensive information on named genera, species, and higher taxonomic groups of Hymenoptera (bees, wasps, etc.) are being computerized for information retrieval purposes, and to publish a revision of the catalog (in Press, 1978).

Invertebrate Zoology Department - Crustacean Collection. One of the first current catalog files started by the museum was for this collection. Major products of the process are catalog cards, geographic cards and specimen

labels. Data for a major accession of 7,900 lots of crustaceans were recently added to the file by conversaion of a file already computerized by another institution.

A related computer file is being built of all collecting-locality information for the ship Albatross, which made substantial collections around the turn of the century. The file will be used to: (1) automatically add locality data to computerized records of specimens from the Albatross, (2) accumulate corrections to the published locality data, (3) distribute updated locality data to those interested, and (4) provide a base for building faunal lists for Albatross localities.

Invertebrate Zoology Department - Mollusk Collection. Current/retrospective catalog files are being built for the cephalopods (squid octupus), Pectens (scallops), tridacnids, and littorinids. A current catalog may be started for the entire collection. Files being built are relevant to current research projects.

Mineral Science Department - Mineral Collection. The entire collection of over 150,000 specimens has been entered in a retrospective index file. The original purpose of the file was to gain better control of this intrinsically valuable collection. Intensive planning is now going on to integrate the file into a computerized registration system, and to greatly increase the accessibility of the file via terminals to better meet daily information needs. Related bibliographic files are maintained by individual scientists.

Mineral Sciences Department - Petrology Collection. A current/retrospective file is being compiled and the entire collection will eventually be entered. As with the mineral collection, the plan is to integrate the file into a computerized registration/information system within the department. Related bibliographic files are maintained by individual scientists.

Mineral Sciences Department - Volcano Reference File. Eruption data for all the active volcanos in the world have been computerized to provide

an information source for scientists. The file aids collection management in that it provides additional verified data on specimens in the collection.

Paleobiology Department - Type Specimens. New type specimens of all fossil animals are entered into a current catalog from which the computer produces catalog cards and specimen labels. Extensive retrospective cataloging has been done for some groups such as the brachiopods and conodonts. Verification of published data is an important objective of the project.

Reference computer files have been built for trilobite literature, and literature mentioning types held in the collection. A special file has also been made for collecting locality information for over a million specimens from the Glass Mountains, Texas. When the specimen data are computerized, the locality data will be automatically added to the records.

Paleobiology Department - Non-Types. Data from selected groups of important non-type specimens have been put in a catalog file. An example is the Burgess Shale Collection which is the subject of intensive research.

Paleobiology Department - Sediments. Data from over 20,000 sediments and core samples have been compiled in a catalog file which is a key part of the management system for the sediment collection. Data are included which will allow sediment-distribution maps to be plotted by the computer. The files has been used to answer questions from private industry. Research files are also being built which contain detailed analyses for some of the samples.

Paleobiology Department - Glass Mountain Species List. A research file was prepared which contains the names and collecting localities of over 750,000 specimens of brachiopods which are the subject of a multi-part publication now nearing completion. The file was primarily intended to produce a listing for the publication, but it will also form the foundation for the catalog file which will be produced after the publication is finished.

Vertebrate Zoology Department - Bird Collection. Data on over 20,000 bird specimens have been compiled in a retrospective catalog file, one of the first computer files in the museum. Early emphasis was on seabirds and some endangered species. Later attention focused on newly accessioned specimens from public health research projects in Cyprus, Ethiopia, and Egypt.

Vertebrate Zoology Department - Bird Names. A reference computer file of all bird names has been built to assist with management of the bird collection. The printout serves as a checklist and data capture form for use in reviewing the contents of the collection and for compiling specimen counts. The list also serves as a finding index to the collection.

Vertebrate Zoology Department - Fish Type Specimens. An index file for all type specimens in the collection is being compiled to verify the inventory, and to provide a listing of types to the scientific community.

Vertebrate Zoology Department - Fish Collection. A current catalog file is maintained for all fish specimens cataloged since 1971. Catalog cards and specimen labels are a product of the computerization process.

Vertebrate Zoology Department - Fish Loans. Index data for all active loans of fish specimens are put in a special computer file. Listings by date and borrower help locate loans which have been outstanding too long. Listings by species name allow curators to locate specimens they wish to examine. Periodic summaries provide information such as number of specimens on loan, number of types on loan, and number of different borrowers.

Vertebrate Zoology Department - Reptiles and Amphibians. A current catalog file has been maintained since 1971, and is one of the oldest of such projects in the museum. An index file was recently prepared for the entire collection from the United States. During the next few months National Museum of Natural History and the National Fish and Wildlife Laboratory of the Department of the Interior will work

together to conduct a physical inventory of the collection and to verify the data

Vertebrate Zoology Department - Mammals Collection. A retrospective/current catalog file is bing built for the entire mammal collection. The computer file is used to print catalogs and specimen labels. The data standards of the project have been adopted by NIRM (National Informational Retrieval Network for Mammalogy) which is striving to build data banks for all mammal specimens in the United States.

Associated files of key interest are: (1) file of marine mammal stranding records and of marine mammal specimens held by other museums; (2) file of data for a huge collection of mammals and associated parasites collected for the U.S. Army in Venezuela; (3) file of data for a large colleciton of mammals from Africa.

Smithsonian Oceanographic Sorting Cente (SOSC) SOSC provides general identification and gross sorting or oceanographic samples such as algae, invertebrates, fish, plankton, geological specimens. These specimens received from several governmental agencies are stored until they are requested by research scientists located throughout the world.

Unidentified Specimen File - This is the largest file and contains records about specimens which have been processed and grossly sorted. Information for each specimen include station data, gross identification, number of specimens, and storage location. This file serves as an inventory and index to the collections.

Identified Biological Specimen File - As specimen lots are studied in detail and identified, records are added to this file.

Bottom Photographs File - A manual file of ocean bottom photographs is maintained. A computerized index file contains data about the geographic location where the bottom photographs were taken and salient biological, physical, and geological features found in the photographs.

Sample Holdings File - This management file

contains a record of all samples received, sorted, and shipped from the oceanographic inventory. In addition to containing station data, this file provides information on the location and processing stage for each sample.

Registry of Tumors In Lower Animals - The Registry of Tumors In Lower Animals has two SELGEM files to assist with their studies. The bibliographic citation index file is an index to all articles dealing with tumors in lower animals. It is indexed by keywords, taxonomic names, locality, and tumor morphology. The second file is a mailing list application with the names and addresses of the researchers in the tumor field.

Center for the Study of Man - The Center for the Study of Man is using SELGEM to assist in the preparation of bibliographic information for the twenty volume Handbook of North American Indians. Hundreds of scholars from all over the world have prepared articles for this comprehensive encyclopedia written from the perspectives of anthropology, history, and linguistics. Full page makeup of the bibliographic section of the Handbook is created by a computer-driven typesetting machine using data collected through the SELGEM system.

2. National Museum of History and Technology (NMHT)

Work is underway to computerize the registration and cataloging systems of the museum. These data files will be used for research, collections management, and inventory control. Very comprehensive records are entered into the SELGEM system. Data elements recorded included: name of object, donor, acquisition history, description of object, and relationship of the object to other objects in the collections.

3. National Air and Space Museum (NASM)

NASM has three SELGEM applications, two in the library and one in collections management.

The library started a periodical citation file in 1973. It is an index to the current literature in the field of air and space. The file

is updated monthly and is used mostly by the museum staff.

The Film Inventory file is an index to the more than 1,000 motion pictures and film strips in the NASM library collections. The films are indexed by subject, title, and location.

An inventory of the collection of the National Air and Space Museum comprising more than 20,000 items is now in progress. These items include armaments, art, awards, crafts, equipment, instruments, models, and engines, etc. About 5,000 items have been inventoried. Indices by subject and department will be prepared upon completion of the inventory. The file will be used for inventory control.

4. Hirshhorn Museum and Sculpture Garden (HMSG)

The Office of the Registrar is responsible for providing information on the physical characteristics of each object in the museum collection, including object class, medium, dimensions, artist's inscription and condition. The complete collection of about 6,400 items has been computerized. A number of practical working tools were developed from this computer file, including artist lists, donor lists, storeroom inventory cards and labels for the photographic archive files.

5. National Collection of Fine Arts (NCFA)

Inventory of American Paintings Executed Before 1914. - This project was begun in 1971 to locate and record paintings from across the nation, with particular concern for those never before recorded. With the aid of individuals and local committees throughout the country, 180,000 paintings have been located and data about them placed into a SELGEM computer file. A guide to the inventory has been published and scholars have been quick to utilize its information. Indices by artist, subject, and owner are prepared annually. This is the largest individual SELGEM file with an accession rate of more than 20,000 records per year.

Slide and Photograph Inventory. This project

has a collection of approximately 20,000 slides and 100,000 photographic negatives. About 37,000 have been placed in the SELGEM system and are indexed by artist, subject, title, and geographic location. Many of the negatives are of the old silver nitrate type which must be rephotographed using modern techniques, new labels are printed from the computer files for these negatives.

Registrar File. - A complete data file on the holding of NCFA is used to provide information on the collections; it numbers about 11,000 items.

6. National Portrait Gallery (NPG)

The National Portrait Gallery Field Survey is an index to portraits held in other museums and private collections of prominent Americans or portraits by prominent American artists. The file contains about 8,000 records with data about the artist, sitter, location, occupation, and medium. The file is used to prepare indices.

The Periodical Portrait Index contains information on approximately 16,000 portraits published in periodicals and catalogs. Indices are prepared by journal, artist, and sitter.

7. Smithsonian Archives (SA)

The Archives is a depository for letters, scientific papers, fieldnotes, and manuscripts. The Archives uses SELGEM for two applications. The Guide to the Smithsonian Archives will be published in the fall of 1978. The Guide describes in general terms over 400 major collections of materials in the Archives. It is being printed from the SELGEM system by a computer typesetting system at the Government Printing Office.

Each collection has a detailed summary description of the collection, listing names, dates, and subject areas of the material, which are called the finding aides. A SELGEM index file to the finding aides contains more than 3,400 records.

8. Conservation Analytical Laboratory (CAL)

The Conservation Analytical Laboratory serves as the central research organization of the Institution in a wide variety of conservation-related fields. A major function of the Laboratory is to disseminate information concerning conservation of museum objects. A computerized file of past treatments is maintained on-line using the SELGEM system. Information stored in this file includes detailed description of the object, the nature of the problem, summary description of the treatment, and the results. A large number of key words by material, decoration, condition, and type of treatment are included for retrieval.

9. Chesapeake Bay Center for Environmental Studies (CBCES)

The Center is studying the Rhode River, Maryland watershed to determine how land management, climate and other factors influence the movements of materials (sediments, nutrients, organic matter, bacteria, and pesticides) from a watershed into an estuary. Automatic sampling stations have been constructed to monitor runoff and rainwater. This data is stored in the SELGEM system for information retrieval and analysis.

VII. Acknowledgments for Development of the SIIR/SELGEM System

In addition to Mr. Reginald Creighton and Mr. James Crockett, Office of Computer Services, credit for creating the SELGEM system goes to programmers of the Office of Computer Services. Mr. William Bradley, Mr. William Handley, Mr. William Lawson, Ms. Penelope Packard, Mr. Victor Ponte, Mr. James Reeves, Mr. Edwin Robinson, Ms. Patricia Tuneski, Ms. Dianne Walker, Mr. Ralph Walker, and Mr. Richard Young, all analyst-programmers wrote or made significant contributions to the thirty or more programs that constitutes the SELGEM system.

Mr. David Bridge of the National Museum of Natural History (NMNH) contributed greatly to the conceptualizing which preceded much of SELGEM development, and he continues to provide much of the management skill which has made the system come alive at the National Museum of Natural History. Credit must go to Dr. Richard C. Cowan, who had the vision to recognize the value of the computer in systematics and established the Automatic Data Processing program at the NMNH, and to Dr. T. Gary Gautier who has managed the program since mid-1973.

Dr. James F. Mello, Assistant Director, National Museum of Natural History (NMNH) helped develop techniques, protocols, and organizational methods. He has communicated ideas and criticisms to keep the system's development in consonance with the requirements of natural history. His imprint on the system is strong.

Mrs. Nancy Parsons patiently typed and retyped the system's documentation. Mr. Dante Piacesi developed the mathematical techniques for geographic retrieval.

The keen and unfailing direction of Dr. Donald F. Squires, past Assistant Director, of the National Museum of Natural History, bolstered our perserverance during the early years of SIIR system development. Mr. S. Dillon Ripley, Secretary of the Smithsonian Institution provided funding for initial SELGEM developments. Mr. Stanley A. Kovy, Director of the Office of Computer Services gave his support and encouragement. Since SELGEM's second year, the NMNH has supported modifications and enhancements to optimize SELGEM's performance on the data of natural history. Other Smithsonian bureaus have also supported SELGEM extensions.

The significant contribution of the late Mr. Kenneth Ebbs was much apprecated.

Dr. Uta Merzbach, Curator, Section of Mathematics, National Museum of History and Technology has contributed her considerable expertise on numerous occasions and we are grateful for the supportive participation of Dr. Melvin H. Jackson of the Division of Transportation. Ms. Katherine Peterson, Collections Registration Specialist, National Museum of History and Technology, contributed many fine ideas soon to be reflected to the system.

We are grateful for the generosity of a number of early pioneers in the National Museum of Natural History who invested their time, energy, intellect, and enthusiasm in a then uncertain data processing odyssey. We thank: Mr. Harold H. Banks, Jr., Museum Specialist, Department of Mineral Sciences; Mr. Colby Allan Child, Museum Specialist, Department of Invertebrate Zoology; Mr. Frederick J. Collier, Collections Manager, Department of Paleobiology; Dr. Mason E. Hale, Jr., Curator, Department of Botany; Dr. Duane W. Hope, Chairman, Department of Invertebrate Zoology; Dr. Raymond B. Manning, Curator, Department of Invertebrate Zoology; Dr. Willaim G. Melson, Curator, Department of Mineral Sciences; Dr. Henry W. Setzer, Curator, Department of Mammals; and Mr. Stanwyn G. Shetler, Associate Curator, Department of Botany.

We thank those who believed in and used the system in its early, more primative state: Ms. Abigail Booth, Coordinator, Bicentennial Inventory of American Paintings; Mr. William Johnson, Registrar, National Collection of Fine Arts; Mr. Joseph Sefekar, Administrator, Hirshhorn Museum and Sculpture Garden; Mrs. Margaret Moore, Technical Information Specialist, formerly of the Hirshhorn Museum of Sculpture Garden; and Nathan Reingold, Editor, Joseph Henry Papers.

VIII. Bibliography

Listed below are some of the more important publications relating to the system's development, application descriptions, and products of the SIIR System and the SELGEM System.

Association of Systematic Collections. SELGEM at the University of Kansas (Paleontology, Entomology, Herbarium, Environmental Questions). Association of Systematic Collections Newsletter, 1(4): 14 (1973).

Ayensu, Edward S. and DeFilipps, Robert A. Endangered and Threatened Plants of the United States. Washington, D. C. XV, 403pp. (1978).

Bain, Alan L. A Proposal for the Computerized Control of Smithsonian Institution Archival Resources. Smithsonian Archives - Xerox Distribution (1973).

_____ Archives Puts Computer System into Effect. The Smithsonian Torch, June 1974.

Bennett, M. Ann. San Juan Valley Archaeological Project SELGEM Data Recording Manual: Ceramic Analysis. Eastern New Mexico University (1974).

_____ Basic Ceramic Analysis. Eastern New Mexico University Contributions in Anthropology, 6(1): 1-183 (1974).

Booth, Abigail V. An Inventory for the Art Researcher. Museum News, 51(4): 37-39 (1972).

_____ The Bicentennial Inventory. American Art Review, September-October 1973.

Bork, Chris J. Museum Uses Smithsonian System: Anthropologists Add Computer to Tools. Arizona Daily Wildcat, p. 5, February 24, 1975.

Bullis, Harvey R., Jr. and Roe, Richard B. A Bionumeric Code Application in Handling Complex and Massive Faunal Data. Systematic Zoology, 16(1): 52-55 (1967).

Chenhall, Robert G. Museum Cataloging in the Computer Age. American Association for State and Local History, Nashville, Tenn. VIII, 261pp. (1975).

Collier, Frederick J. Catalog of Type Specimens of Invertebrate Fossils: Conodonta. Smithsonian Contributions to Paleobiology, 9:1-256 (1971).

Creighton, Reginald. An Information Storage and Retrieval System for Biological and Geological Data: Design Considerations. In: Symposium sobre Problemas de Informacion en en Ciencias Naturales, Mexico D.F. Dicembre 12-18, 1967.

_____. The SELGEM System for Information Management: An Introduction. (Brochure), November 1974.

_____, and Crockett, James J. SELGEM: A System for Collection Management. Smithsonian Institution Information Systems Innovations, 2(3): 1-24 (1971).

_____, and King, Richard. The Smithsonian Institution Information Retrieval (SIIR) System for Biological and Petrological Data. Smithsonian Institution Information Systems Innovations, 1(1): 1-25 (1969).

_____, and King, Richard. The Smithsonian Institution Information Retrieval (SIIR) System . In: Proceedings of the Sixth Annual National Colloquium on Information Retrieval, ed. L. Schultz. Philadelphia: Medical Documentation Service, The College of Pysicians of Philadelphia. pp. 31-50.

_____, and Packard, Penelope. Museum Exchange of System's Help (MESH). The SELGEM Newsletter, No. 1, April 1973.

_____, and Packard, Penelope. Computer Assisted Information Management: Getting Oriented. Smithsonian Institution Procedures in Computer Sciences, 1(2): 1-19 (1974).

_____, and Packard, Penelope. Museum Exchange of System's Help (MESH). The SELGEM Newsletter, No. 2, April 1974.

_____, and Packard, Penelope. Museum Exchange of System's Help (MESH). The SELGEM Newsletter, No. 3, January 1976.

_____; Packard, Penelope; and Crockett, James. File Building: A Manual for Data Banking via SELGEM. Smithsonian Institution Procedures in Computer Sciences, 1(2): 1-102 (1973).

_____; Packard, Penelope; and Linn, Holley. SELGEM Retrieval: A general description. Smithsonian Institution Procedures in Computer Sciences, I(1): 1:38 (1972).

_____; Packard, Penelope; and Linn, Holley. SELGEM RETRIEVAL: A General Description. In: Cuadernos de Historia Economica de Cataluna, IX, Barcelona, Junio de 1973.

_____, Parsons, Margaret. Origins and goals of MESH. Curator, 20(1): 23-31 (1977).

_____, and Robinson, Edwin. Computer assisted information management: Input and file maintenance. Smithsonian Institution Procedures in Computer Sciences, 1(4): i-v, 1-36 (1975).

Crompton, A.W. The Present and Future Course of our Museum. Museum News, 46(5): 35-37 (1968).

Crovello, Theodore J., and MacDonald, Robert D. Index of EDP-IR Projects in Systematics. Taxon, 19(1): 63-76 (1970).

Cutbill, John L. ed. Data processing in biology and geology. The Systematics Association Special Volume No. 3. New York, New York. xv, 346pp. (1971).

Darnell, Rezneat M.; Lemon, Paul C.;Neuhold, John M.; and Ray, G. Carleton. Natural areas and their role in land and water resources preservation. Final Report to the National Science Foundation. viii, 286pp. (1974).

Ernst and Ernst. SELGEN - PACKAGE DESCRIPTION - Museum Data Bank Coordinating Committee, September 1973. pp. 151-243.

Erwin, La Verne J. Magarian. Application of a computerized general purpose information management system (SELGEM) to natural history research data bank (Coleoptera: Carabidae). Coleopterists Bulletin, 30(1): 1-32 (1976).

Field Museum of Natural History. Cataloguing Comes of Age. A SELGEM information package, Chicago, Ill. (1972).

Gaines, Sylvia W. MESH. Newsletter of Computer Archaeology, p. 8 (1973).

_____. Computer Applications in Archaeology at University of Arizona. Newsletter of Computer Archaeology, 9(3): 4-6 (1974).

Galler, Sidney R.; Oliver, James A.; Roberts, H. Radclyffe; Friedmann, Herbert; and Squires, Donald F. Museums Today. Science, 161(3841), pp. 548-551 (1968).

Gautier, T. Gary. A Report on the Use of SELGEM at the National Museum of Natural History. Association of Systematic Collections Newsletter, 2(1): 12-17 (1974).

Hale, Mason E., and Shetler, Stanwyn G. Botanical type specimen register, part 2. Washington, D. C. Flora North America Report, 76: x, 1-351 (=178pp.) 1974.

Hattery, Lowell H., ed. The SELGEM System at Smithsonian. Information Retrieval and Library Automation, 9(9):12 (1974).

Hodgson, Frank. The Collection and Storage of Genetic Samples from Rare Animals. Association of Systematic Collections Newsletter, II(3) (1974).

Hope, W. D., and Murphy, D. G. A Taxonomic Hierarchy and Checklist of the Genera and Higher Taxa of Marine Nematodes. Smithsonian Contributions to Zoology, 137: 1-101 (1972).

Hubaux, A. Geological Data Files Results of an International Inquiry. Published by International Union of Geological Sciences, December 1971, DOC 31.

_____. A New Geological Tool-The Data. Earth Sciences Reviews, 9: 159-196, Elsevier Scientific Publishing Company, Amsterdam, The Netherlands (1973).

Hurd, Paul D., Jr. Annotated catalog of the carpenter bees of the western hemisphere. Washington, D. C. vi, 120pp. (in press) 1978.

International Business Machines Corporation, Data Processing Division. Computers in the museum. White Plains, New York 69pp. (1973).

Irwin-Williams, Cynthia, and Clark, Paddy. The Development of Data Storage and Retrieval Systems in Archaeology: The San Juan Valley Archaeology Program. Newsletter of Computer Archaeology, 8(4): 4-5 (1973).

Klausen, Ivar. Informasjonssystemet SELGEM ved Smithsonian Institution. Reiserapport/foredrag holdt ved NORD-FORSK's symposium: Strukturering og kodifisering av stedsbundne maledata fra vannmiljo, Kongsberg, 31.10-2.11 (1973).

_____. SELGEM: Et Automatisert Informasjonssystem Universitetet i Trondheim. Det Ngl. Norske Videnskabers Selskab, Museet Erling Skakkesgt 47b, N-7000 Trondheim, Norway. (1974).

Krombein, Karl V.; Hurd, Paul D., Jr.; Smith, David R.; and Burks, B.D., eds. Catalog of hymenoptera in America north of Mexico. Washington D.C. 3 vols., 3000+ pages (in press) 1978.

_____; Mello, James F.; and Crockett, James J. The North American Hymenoptera Catalog: A Pioneering Effort in Computerized Publication. Bulletin of the Entomological Society of America, 20(1): 24-29 (1974).

Lemon, Paul C. AIBS Conservation of Ecosystems Program. Bio-Science, 24(4):249 (1974).

Manire, Larry. The SELGEM System. Newsletter of Computer Archaeology, 8(4): 2-3 (1973).

_____. Current Status of Computerization Activities at the Arizona State Museum. February 1974 (Xerox Copy).

_____. Methods for the Computerization of the Canyon del Muerto Site Survey - (CDMS). Arizona Archaeological Center of the National Park Service and the Arizona State Museum, University of Arizona, 1-49 (Xerox Copy). 1975.

Manning, Raymond B. A Computer-Generated Catalogue of Types: A By-product of Data Processing in Museums. Curator, 12(2): 134-138 (1969).

_____. Automation in Museum Collections. Proceedings of the Biological Society of Washington, 82:671-686. (1969).

Mello, James F. Paleontologic Data Storage and Retrieval. Proceedings of the North America Paleontologic Convention Part B, pp. 57-71 (1969).

_____. Computer Revolution in Systematics. Taxon, 23(1): 21-22 (1973).

_____. Brenan, J.P.; Rose, R.; and Williams, J.T. eds. The use of the SELGEM System in Support of Systematics. In: Computers in Botanical Collections. New York, pp. 125-140 (1975).

_____, and Collier, Fredrick J. New Proceedings in Recording Specimen-Related Data on Fossils. Journal of Paleontology, 46(5): 776-777 (1974).

Nagle, Christopher. Report on meeting for computer data banking in anthropology museums. Newsletter of Computer Archaeology. September 1976.

Nall, Lexi. Computerized Storage and Retrieval System: STENGEM. Stanford University, Department of Anthropology, Enthnographic and Archaeological Collections (1975).

Neuner, A. M. SELGEM manual. Association of Systematic Collections. Lawrence, Kansas. vii, 104pp. (1976).

_____. SELGEM workbook. Association of Systematic Collections. Lawrence, Kansas, viii, 50pp. (1976).

Ortner, Donald J., and Monahan, Marguerite. The paleopathology program at the Smithsonian Institution. Paleopathology Newsletter, 10: 7-8 (1975).

Piacesi, Dante, and Creighton, Reginald A. An Approach to the Geography Problem in Museums, in Proceedings of the Sixth Annual National Colloquium on Information Retrieval, ed. L. Schultz, Philadelphia: Medical Documentation Service, The College of Physicians of Philadelphia, pp. 441-456 (1969).

Purdy, Virginia C., and Crockett, James J. Automation of the catalogue of American Portraits. Smithsonian Institution Information Systems Innovations, 1(2): 1-9 (1969).

Radford, Albert E.; Dickson, William C.; and Massey, Jimmy R.; Ritchie, C.R.; and collaborators. Information Systems and Data Banking. In: Vascular Plant Systematics, Harper and Row, 912 pp.

Simmons, Keith L., and Landrum, Betty J. Sea Floor Photographs. Antarctic Journal of the United States,

8(3): 128-133 (1973).

Shetler, Stanwyn G. The Flora North America System. Proceedings of the 1973 Symposium at Kew, NATO Ecosciences Pannel (1973).

_____, et al. An introduction to the botanical Type Specimens Register. Smithsonian Contributions to Botany, 12:V 1-186 (1973).

Smithsonian Institution. Hymenoptera Catalog is Computerized. Smithsonian Institution Research Reports, 2:3 (1972).

_____. Smithsonian Scientists Gather Computerized Data on Volcanic Activity. In: Smithsonian Institution Research Reports, Number 1 (1972).

_____. SI Computer System May Become Standard for Museum Usage. Smithsonian Institution Research Reports, 4: 3,7 (1973).

_____, Department of Paleobiology. Fossil Marine Mammals in the Collections of the United States National Museum of Natural History, Washington, D. C., 173pp. (1971).

Squires, Donald F. Data Processing and Museum Collections: A Problem for the Present. Curator, 9(3): 216-227. (1966).

_____. Collections and the Computer. Bioscience, 18(10): 973-974 (1968).

_____. Schizophrenia: The Plight of the Natural History Curator. Museum News, 48(7): 18-21 (1969).

_____. An Information Storage and Retrieval System for Biological and Geological Data. Curator, 8(1): 43-62 (1970).

_____. Implications of data processing for musuems. In: The Systematic Association Special Volume No.3. Data Processing in biology and geology. ed. J.L. Cutbill. pp. 235-253 (1971).

Sutton, John. SELGEM at Texas Tech University. Association of Systematic Collections Newsletter, 1(2) (1973).

Sturtevant, William C. Gen. Ed. Handbook of North

American Indians: Vol. 8, California. Washington, D.C. xv, 800pp. (1978).

_____. Handbook of North American Indians: Vol. 15, Northeast. Washington, D. C. (in press) 1978.

Tamplin, Morgan J. Computers and Anthropology at the University of Arizona, University of Arizona, Tuscon, April 1973.

Thompson, Raymond H. A..GEM of a System. Museum News, 52(1): 8 (1973).

United States House of Representatives. Report on endangered and threatened plant species of the United States, presented to the Congress of the United States of America by the Secretary, Smithsonian Institution. House Document 94-51. Washington, D. C. iv, 200pp. (1975).

Vance, David. Smithsonian Institution. In: Computers in the Museum, IBM Publication, pp. 47-48 (1973).

Warren, Nancy; Klausner, Stephanie; Meek, Kathy; and Reiley, Anita. A SELGEM Procedural Manual for the Laboratory of Anthropology Archaeological Inventory. 12 pp. (Xerox Copy) 1974.

Weisbrich, Alfred L. NASM DRS Presentation Report. National Air and Space Museum, Smithsonian Institution, Washington, D. C. (1973).

Wing, Elizabeth S. Early Domestic Animals. Domestikationsforschung und Geschichte der Hustiere. (Proceedings of a Symposium), Budapest, Hungary (1974).